

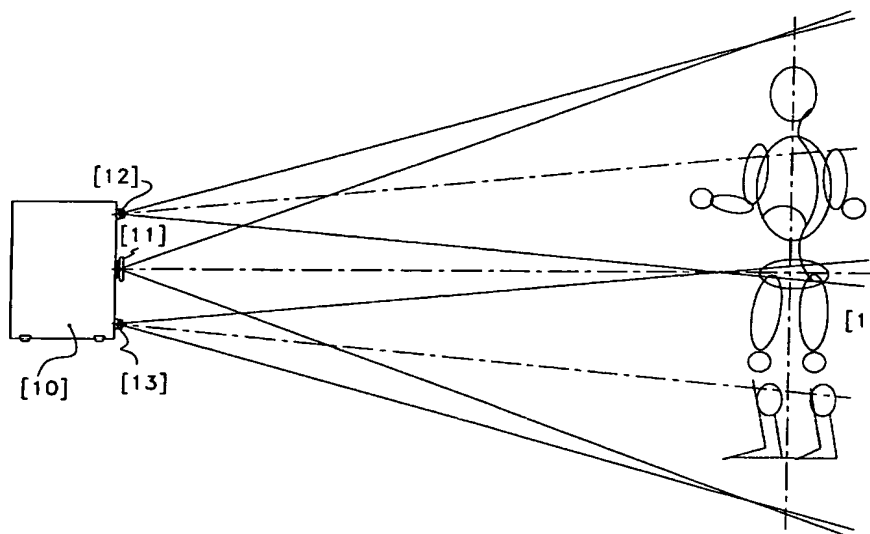


(71) INSPECK INC., CA

(51) Int.Cl.<sup>6</sup> G01B 11/24, A61B 5/107

(54) **NUMERISEUR OPTIQUE 3D DE TOUT LE CORPS HUMAIN**

(54) **OPTICAL FULL HUMAN BODY 3D DIGITIZER**



## OPTICAL FULL HUMAN BODY 3D DIGITIZER

### FIELD OF THE INVENTION

The present invention relates to an optical full human body 3D digitizer that has numerous industrial applications, for example color non-contact optical 3D digitizing, computer assisted 3D vision, human body digitizing, computer animation and computer graphics, electronic gaming, 3D electronic achieving, 3D web, reverse engineering and medical 3D imaging.

### BACKGROUND

3D digitizing, particularly non-contact optical 3D digitizing techniques, became commercially available during recent years. Most of these techniques are based on the principle of optical triangulation. Despite the fact that passive optical triangulation (stereo vision) has been studied and used for many years for photogrametric measurements, the active optical triangulation technique (particularly laser scanning technique) gained popularity because of its robustness and simplicity to process obtained data using a computer. Most of the systems based on the active optical triangulation principle were developed for industrial applications, such as robotic assembly, robot guidance, industrial inspection, reverse engineering, etc.

A laser beam or a laser stripe is projected on a 3D surface of an object, scattering the laser beam or laser stripe on the surface. It is measured using a photo-electronic device. A signal can be measured indicating the position (usually the depth) of the measuring point. In most cases, the basic measurements are either a point or a section profile. A mechanical or optical scanning device is usually used to provide a frame of 3D measurement. For industrial applications, mechanical scanning can be accomplished by the mechanism on which the digitizing device is mounted, such as a robot or a conveyer. The scanning process is a sequential data acquisition process and takes relatively longer time to scan a surface. During the scanning, the object should be kept immobilized; this is a major problem when scanning a live being. Different techniques, such as the projection of multiple stripes, laser line scanning during one video frame and high speed scanning, have been developed. These approaches are either too

expensive to realize, or their sampling rate is still too low compared to 2D digital imaging.

A laser beam is a monochromatic light source. One single monochromatic laser beam can not provide full color information of the measured surface. On the other hand, a number of today's 3D applications including computer animation, electronic games, 3D Web, 3D archiving and 3D medical imaging require information on color texture which contribute to most of the visual effects. In order to measure the color texture of a surface, a 3D digitizing system based on a laser scanning principle has to use multiple laser sources (blue, green and red lasers) or use a second camera to get color data. The first solution is very difficult to be implemented and is also very expensive. The second can suffer from problems of misalignment between 3D geometric data and color texture data because they are not captured from the same angle of the view.

When digitizing a full human body, the required ratio between height and width of the measured zone should be 2 to 3 over 1. A system based on laser scanning is more flexible to provide a desired ratio, but its acquisition speed is too slow. Any other systems using frame capturing of a CCD camera are limited by the geometric form of the sensor. Most of commercially available CCD sensors have an aspect ratio equal either to 4/3 or to 1. If such a sensor is used to cover a human body possibly higher than 2 meters, the resulting lateral resolution would be very low. At the same time, a many of the pixels are not useful for a measurement.

## SUMMARY

An object of the invention is to provide an optical full human body 3D digitizer that addresses the above drawbacks of existing optical 3D digitizers.

A subsidiary object of the invention is to provide a reliable solution for a cost effective system. .

According to the present invention, there is provided an optical full human body 3D digitizer comprising two standard color (color version) or monochromatic (B/W version) cameras and one single white light projector. The two cameras are set in a way that over all the depth of the measurement, their captured images are always overlapped. The two images from the two cameras can be merged to form

one single image. The aspect ratio of the combined image varies between 2 to 3 over 1. One or a few combined images will be required to provide 3D measurement of one view of a human body. The acquisition time of one view requires a fraction of a second using commercially available standard cameras and frame grabbers. The cameras used for the measurement of 3D geometry provide also the capturing of color or gray scale texture, depending on the cameras. The mapping of texture on top of 3D geometry is automatically ensured by the nature of the data acquisitions.

In addition, solutions for different technical features related to the above digitizer are proposed. Two approaches have been developed for 3D coordinate measurements. A first one uses one video frame containing a projected fringe pattern and a second one requires a few video images which also contain a projected fringe pattern. An image processing technique based on the analysis of mechanical interference pattern provides the 3D coordinate data for each image pixel. An encoding technique is applied to ensure the conversion of the measurement in computer units to real physical parameters. A defocusing optical element is provided to remove fringe patterns from the image and provide a uniform illumination. In order to keep light intensities similar for two images grabbed using two separate cameras, a procedure for light intensity adjustment is implemented, which uses the average light intensity measured on the overlapped area of the two images. The data obtained from each camera should be calibrated in a common coordinate for both cameras. Both 3D geometric data and texture data acquired by the two cameras are merged to form one single 3D model with one single texture image.

The present invention provides ways to create a complete model of a human body using single or multiple optical full human body 3D digitizers.

#### BRIEF DESCRIPTION OF THE DRAWINGS

A detailed description of preferred embodiments will be given herein below with reference to the following drawings, in which like numbers refer to like elements:

Figure 1 is a schematic diagram showing a color and/or monochromatic optical full human body 3D digitizer and a person to be digitized, according to the invention;

Figure 2 is a schematic diagram showing a rotational wheel containing a defocusing device, a shutter and filters according to the invention;

Figure 3 is a schematic diagram showing projected pattern and encoding points according to the invention;

Figure 4 is a schematic diagram showing a color and/or monochromatic optical full human body 3D digitizer according to the invention;

Figure 5 is a block diagram of a control circuit for the digitizer according to the invention;

Figure 6 is a flow chart showing a control sequence 3D acquisition, according to the invention; and

Figure 7A-C are schematic diagrams showing a fringe pattern displacement, according to the invention.

#### IDENTIFICATION OF THE COMPONENTS

The following is a list of the reference numerals, along with the names of the corresponding elements, that are used in the appended drawings and in the description.

- (01) Person to be digitized.
- (10) Color or/and monochrome optical full human body 3D digitizer
- (11) Projection lens
- (12) Camera 1
- (12a) Lens of camera 1
- (13) Camera 2
- (13a) Lens of camera 2
- (15) SPI
- (16) CPU
- (17) RAM& ROM Memory
- (18) Input / Output
- (19) Motor driver board
- (24) Lamp driver board

- (25) White light source projector
- (26a) Power supply (Cameras)
- (26b) Power supply (Motors)
- (26c) Power supply (Control board)
- 5 (27) Potentiometer
- (30) Main control board
- (31) Fringe pattern and fringe pattern positioning device
- (32) Rotational wheel
- (33) Zero sensor position fringe pattern positioning device
- 10 (34) Zero sensor position filter wheel
- (37) Defocusing device
- (38) Filters
- (39) Shutter
- (40) Computer
- 15 (41) Frame Grabber
- (50) Camera cable connectors
- (51) Control cable connector
- (52) Power cable connector
- (53) Cooling fan
- 20 (60) Projected pattern
- (61) Encoding points

#### DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Referring to Figures 1 and 4, in brief, the optical full human body 3D digitizer (10) according to the invention has two standard color (color version) or  
 25 monochromatic (B/W version) cameras (12,13) and one single white light projector (25). The two cameras (12,13) are set in a way that over all the depth of the measurement, their captured images always overlap. The two images from the two cameras (12,13) are merged to form one single image. The aspect ratio of the combined image varies between 2 to 3 over 1. An image of 640X480 pixels can be  
 30 grabbed using one standard NTSC camera. Although there is an overlapped area between the two video images captured by the two cameras (12,13), the final merged image can still keep at least 1100 to 1200X480 pixels. These image pixels

are distributed over a field of view corresponding to the form of a human body. So, there are no wasted image pixels. In addition, this number of sampling over the field of view ensures a very reasonable lateral resolution for final 3D measurement. The number of image pixels captured by each camera is limited  
5 either by industrial standards (for example, NTSC or PAL) or by the manufacturing cost of the camera. This limitation does not apply to the projected pattern (60) as shown in Figure 3. In fact, the only limit for a projected pattern (60) (a film, for example) is the optical resolution of the film and projection optics. It is not uncommon to obtain a resolution 50 to 100 lines per mm on a pattern to be  
10 projected which may have a size of 35mm×25mm. So it is evident that one projected pattern can easily provide the necessary image information for the area covered by the two cameras (12,13). The major advantage of using one single projector instead of two is to avoid the cross-talking results from simultaneous images captured and two fringe patterns if two projectors are used.

15 Referring to Figure 3, two approaches for 3D coordinate measurements are provided. The first one uses one video frame containing a projected fringe pattern (60) and the second one requires a few video images which also contain a projected fringe pattern (60). An image processing technique based on the analysis of mechanical interference pattern provides the 3D coordinate data for  
20 each image pixel. So one or more combined images are required to provide 3D measurement of one view of a human body (1). The acquisition time of one view requires a fraction of a second using commercially available standard cameras (12,13) and frame grabbers (41), as shown in Figure 5. The necessary acquisition time of this system is much shorter than most of existing techniques based on  
25 laser scanning principles and many more data points can be measured on a person who does not need special training to be kept immobilized for several seconds.

The cameras (12,13) used for the measurement of 3D geometry provide directly the capturing of color or gray scale texture (108), as shown in Figure 8. In  
30 order to ensure a uniform illumination during the capturing of texture, a defocusing optical (37) element is introduced, which removes the fringe pattern (60) from the image. Since the same image pixel of the camera measures the 3D geometry and texture data (108) of a point on a 3D surface, the texturing mapping on top of 3D

geometry is automatically ensured by the nature of this data acquisition. The 3D digitizer based on this invention does not need two image sensors to separately measure 3D geometry and texture and avoids the problem of misalignment occurring with laser scanning systems. In addition, since two cameras (12,13) are used to cover a large field of view, it is important to keep light intensities similar for two images grabbed using two separate cameras (12,13). A procedure for light intensity adjustment is implanted which uses average light intensity measured on the overlapped area of the two images.

The data of 3D coordinates obtained from each camera (12,13) should be calibrated in a common coordinate for both cameras. An encoding technique is applied. A set of encoding points (61) is generated by a projected pattern (60) and the absolute positions of these points can be determined once they are measured by the cameras (12,13). In fact, a function describing the absolute positions of the encoding points and their measured position on the photo sensitive area of the cameras can be defined experimentally after a digitizer is assembled. Each camera (12,13) should capture at least one encoding point. The encoding point (61) ensures first the conversion of the measurement in computer unit to real physical parameters for the whole surface and indicates the geometric relation of the 3D images measured by each of the two cameras (12,13). A fine tuning procedure using the 3D data on the overlapped surface gives final adjustment to the positions of the two images. In addition, both 3D geometric data (106) and texture data (108) acquired by two cameras (12,13) should be merged to form one single 3D model with one single texture image.

Finally, two approaches allow to create a complete model of a human body using one or multiple optical full human body 3D digitizers. When one digitizer is used to capture multiple views of a human body, one has to rotate the person to be digitized or to rotate the digitizer around the person so that each necessary view can be measured. It is important to overlap each sequential measurement of the surfaces. It is unnecessary to know the exact position of each acquisition. The texture and geometric data on the overlapped area will be used to ensure the registration of each partial model. In order to reduce total acquisition time, it is possible to use a number of 3D digitizers mounted in a fixed space. Four to six digitizers are usually needed to minimize uncovered surfaces. When this approach



is used, the procedure for the registration becomes more simple because the positions of each view are well known.

Referring to Figure 4, there is shown an embodiment of a color and/or monochromatic optical full human body 3D digitizer (10) according to the invention. The pattern (60) shown in Figure 3 is illuminated by the white light source projector (25) and projected by the projection lens (11). A cooling system (53), can be used to keep the system at an accepted temperature when necessary. The deformed pattern due to the relief of a person's body is captured by two cameras (12,13) through the lenses of the cameras (12a,13a). It is important to keep some overlap between two fields covered by two cameras (12,13). The two cameras (12,13) are powered by a power supply (26a).

Referring to Figures 4 and 5, the video signals are sent to a frame grabber (41) in a PC computer (40) by two video cables connected to the connectors (50). A number of patterns can be captured for different pattern positions obtained by shifting the projected pattern (60) with an integrated fringe pattern positioning device (31). The shifting of pattern and illumination lighting are controlled via a main control board (30) by the PC computer (40) through a cable connected to a connector (51). The intensity of the lighting can also be controlled by a lamp driver board (24). The rotational wheel (32) including the components like the shutters (39), the defocusing devices (37), and the filters (38) shown in Figure 3, are driven by the motor driver board (19) and these components can be used to provide different functions of the image acquisition.

For every recording sequence, a set of video images is processed into a software application to retrieve the 3D information from the structured images, along with the texture information.

Different algorithms can be used for retrieving the shape and texture information from the video images using a projected structured light. In most cases, a hybrid algorithm based on interferometric techniques and active triangulation with different assumptions, is used. This basic algorithm can be applied to 3 frames (and more) recording with temporal phase shifting.

Each of the phase shifting algorithms will allow for extracting the 3D information for every pixel of the video image acquired during the recording. Linked with a very low acquisition time (few video images), this characteristic is a

major advantage compared to other optical 3D measurement techniques (see Figure 7).

From one sequence of video images with structured light projection, the 3D data set can be extracted for every pixel of the camera, with or without the texture image. As the structured light was projected on the object to measure with known position values, a set of equations can be built to represent the phase function of the object for that particular recording. This phase function is defined through the algorithm developed for every sampled point, the pixels of the video image. This algorithm allows for the retrieval of the phase function in a limited range, from 0 to  $2\pi$  radians.

An example of a phase function is:

$$\varphi(i, j) = F(I_1(i, j), I_2(i, j), \dots, I_N(i, j)) = [0, 2\pi] \quad (1)$$

From the characteristics of the CCD camera (12,13) and frame grabber (41), the precision for each digitized level in intensity can be extracted. From this precision range, a local possible variation of the phase function can be obtained. A special filter is applied to the multiple possible phase functions to get the most probable phase value for every pixel from its precision range and its position to its neighborhood and their precision range. It is important to note that the phase modification done during the filtering gives a value within the precision range for the phase function of every pixel: this point is important as the new filtered phase function is not an approximation of the phase function but a noise reduced phase function within the precision range for every pixel. In that case, the new filtered phase function is at least as good as the initial phase function, and in most cases, better. Also, this special filter is not affected by the color of the object to measure, as long as the light coming through the camera lies in the dynamic range of the imaging camera's sensor.

The following equations depict such a special phase filtering:

$$\varphi(i, j) = F(I_1(i, j), I_2(i, j), \dots, I_N(i, j)) + \Delta F(\Delta I_1(i, j), \Delta I_2(i, j), \dots, \Delta I_N(i, j)) \quad (2)$$

$$\varphi(i, j) = \text{range of possible values with different weight} \quad (3)$$

$$\text{Filter}\{\varphi(i, j)\} = \text{choice of most possible value within this range for every } (i, j) \quad (4)$$

Once the phase function is defined for every pixel of the image, a phase unwrapping algorithm developed for speed and robustness, should be used. The

algorithm is preferably optimized for speed for use on a mid-range PC (or any computer type) with a decent processing time but also with some robustness against noisy pixels, bad lightning conditions, bad surface conditions, or errors in the measurement.

5 An example of an unwrapping function is:

$$P\{\varphi(i, j)\} = \vartheta(i, j) + \varphi(i, j) \quad (5)$$

As the phase function is first expressed in a modulo  $2\pi$ , when the value of a complete discontinuity on a 3D surface is bigger than  $1\pi$ , it is possible to get errors in the phase unwrapping algorithm, with an integral number of order, i.e. a multiple  
10 of  $2\pi$  in the phase. Before going through the data calibration process, it is necessary to correct these measurement discontinuities with a discontinuity tracking algorithm. After the application of this algorithm, the unwrapped phase function is ready to be processed with the conversion to real unit algorithm.

A set of reference points is encoded into the structured light projection on  
15 the object to measure. These encoding points allows for the absolute, but low resolution, measurement of the 3D coordinates at the corresponding points on a given surface of the object. On the other hand, the interference phase function provides the higher resolution position for every points of the image covering the object.

20 The phase function, once unwrapped, is now ready to be converted into real coordinates, like millimeters. From a calibration table created based on each digitizer, the measurement of every sampled point is converted to geometric units. This process corrects any distortion in the 3D measurement.

The process is depicted as follows:

$$25 \quad C(\vartheta(i, j) + \varphi(i, j)) = \text{conversion from } (i, j) \text{ to mesh } (x, y, z) \text{ in mm} \quad (6)$$

With the video images processed for retrieving the 3D data set for the object to measure, a numerical algorithm can also be applied to calculate the texture image from the structured light projection. The texture image is a video image showing in gray scale or in color the optical intensity of the measured  
30 surface. This texture image can be mapped onto the 3D data set to increase the realism of the numerical 3D model obtained.

A special video image obtained with the use of a defocusing device can also be used. In that case, this supplementary image shows the color of the object without the structured light projected on it. This technique avoids the residual structured light pattern caused by slight movement of the object during the digitizing process.

The texture image building is depicted as follows:

$$T(i, j) = F_T(I_1(i, j), I_2(i, j), \dots, I_N(i, j)) \text{ or } F_T(I_{N+1}(i, j)) \quad (7)$$

Since the optical full human body 3D digitizer (10) consists of two cameras (12,13) and one light projector (25), one of the major problems to be dealt with is to balance the colors and the light intensity between the two cameras (12,13). This must be done to ensure there is no discontinuity in the combined texture frame. There are two operations to be done: correctly balance the iris apertures (manual adjustment during raw image acquisition) and automatic color balance done at processing time (even in the case of a perfect manual mean intensity balance, it is important to automatically balance the colors because the cameras (12,13) will not be perfectly calibrated, and even if they were so, their responses would probably not be the same at different light intensity levels).

The algorithms use the overlapped regions to compute the mean intensity and the color ratios. Since these two regions cover the same area, they give the scaling factors required to correct the image from each camera.

A manual iris aperture adjustment algorithm is provided. During live video stage, the mean intensities are measured on both overlapped regions after each acquisition and their values are written on the screen beside an indicator that is red when the difference is larger than a given tolerance and that turns green when the mean intensities match within the same given tolerance. The adjustment is thus done in real time by the operator by opening or closing the iris on the cameras.

An automatic color balance algorithm is also provided. This balance may be enabled or disabled by the operator. The algorithm proceeds as follows.

a) The program computes the mean values for each R,G and B channels on both overlapped regions;

b) It then computes a mean value and a scaling factor for each one of them (so two scaling factors in R are obtained, etc., one being slightly larger than 1.0 and one being slightly smaller than 1.0);

c) The scaling factors are applied on each color channel of each image to make them match together. Because the mean light intensity is assumed to be balanced (even roughly) between both images, the color "re-scaling" should not suffer too severely from a mismatching due to a difference of illumination.

Referring to Figure 5, the control system according to the invention is composed of three electronic modules: the main control board (30), the motor driver board (19) and the lamp driver board (24). Each control board is powered by a different power supply (26).

The main control board (30) receives and transmits commands from and to the computer (10) via a RS-232C (43) asynchronous serial bus. This board has a CPU (16), a SPI (Serial Peripheral Interface) (15), an input/output interface (18), counters (19) and RAM (17a) and PROM (17b) memory for data and program.

The commands that are recognized by the board are :

- Positioning, setting the velocity and setting the acceleration of the motor that pulls and pushes on the fringe pattern positioning device (31);
- Resetting the position of the fringe pattern positioning device (31);
- Positioning, setting the velocity and setting the acceleration of the motor that turns the filter wheel (32);
- Resetting the position of the filter wheel (32);
- Turning the projector's light (25) on and off.

The control board (30) transmits direction (35a) and start (35a) to each motors. Positioning, velocity and acceleration information are sent to counters (19). These counters (19) generate a square pulse at a variable frequency. The position is equal to the number of rising edges sent to the motor control board, the velocity to the frequency these edges are sent, and the acceleration of the rate the velocity changes.

For the reset, the fringe pattern positioning device (31) and the filter wheel (32) have position sensors (33,34) that transmit an electrical signal to the CPU (16) via the PIO (Peripheral Input/Output) (18).

The lamp control board (24) receives the on/off signal via the main control board (30). The intensity of the lamp can be manually adjusted with a potentiometer (27).

5 The cameras (12) (13) are controlled by a frame grabber (11) directly plugged in the computer bus (42).

In order to use the optical full human body 3D digitizer (10) on a practical basis, the different views obtained by the invention must be assembled together to form a complete 3D model. This implies registering the views together, merging the 3D points, and merging the texture to form a unique model containing the  
10 information of the complete body. Details of the treatment on the data implied in each of these steps is given hereinbelow.

The first step consists of putting all of the 3D models in the same reference frame; this is called the registration. To do this, one view and its reference frame is arbitrarily chosen to be the reference coordinate in which all of the views of the  
15 digitized human body are intended to be represented. This model is called the fixed model. At the end of this process, all the 3D points of the models will be represented in the reference frame of the fixed model.

By tagging three texture points belonging to both the fixed model and a second model, it will be possible to put this second model in the world of the fixed  
20 one. Of course, these two models must have some overlapping points and texture. A first approximation of the transformation needed to put the second model in the right place is computed by superposing the three tagged points. As many tagged points as desired can be used to compute this initial solution, but a minimum of three is required: the better the initial solution, the faster and more accurate is the  
25 final solution. Then an algorithm is used to minimize the distance between all of the overlapping regions of the two views. The second model being now in the reference frame of the fixed one, the other model can be tagged to this one, and the process just described can be repeated. This procedure must be repeated until  
30 all of the models are in the same world. As some errors can be distributed along in this iterative process (mostly caused by the noise coming from the acquisition procedure), an algorithm that minimize the total distance between all of the models is ultimately used. The final result is the views being placed in a fashion representative of the human body that is digitized.

The second step consists of merging all the points from the separate views to form one 3D model. This can be done either by a cylindrical projection or by a cubical projection. To merge a human body, the use of a cubical projection is used, as some parts of the body are not fully visible from a cylindrical point of view. Some weighted averaging is used when more than one point represents the univocal surface to be obtained. The final number of 3D points representing the final model can be determined at this step.

Finally, the texture of the different views are merged together and the resulting 2D bitmap are attached to the 3D points of the model. This can be either done by a cylindrical or a cubical projection. Again, cubical projection is used, for the same reason mentioned for the merging of the 3D points. As some parts of the texture of the individual models overlap, a weighted average is used to get the final texture. The weight is representative of the reliability of each of the 2D texture points, determined by the angle between the model's normal and the camera during the capturing of the 3D points and 2D texture. The final model is a polygonal mesh. If needed, the number of points representing the surface can be reduced to an appropriate value asked by the intended use of the full human body 3D model.

While embodiments of this invention have been illustrated in the accompanying drawings and described above, it will be evident to those skilled in the art that changes and modifications may be made therein without departing from the essence of this invention.

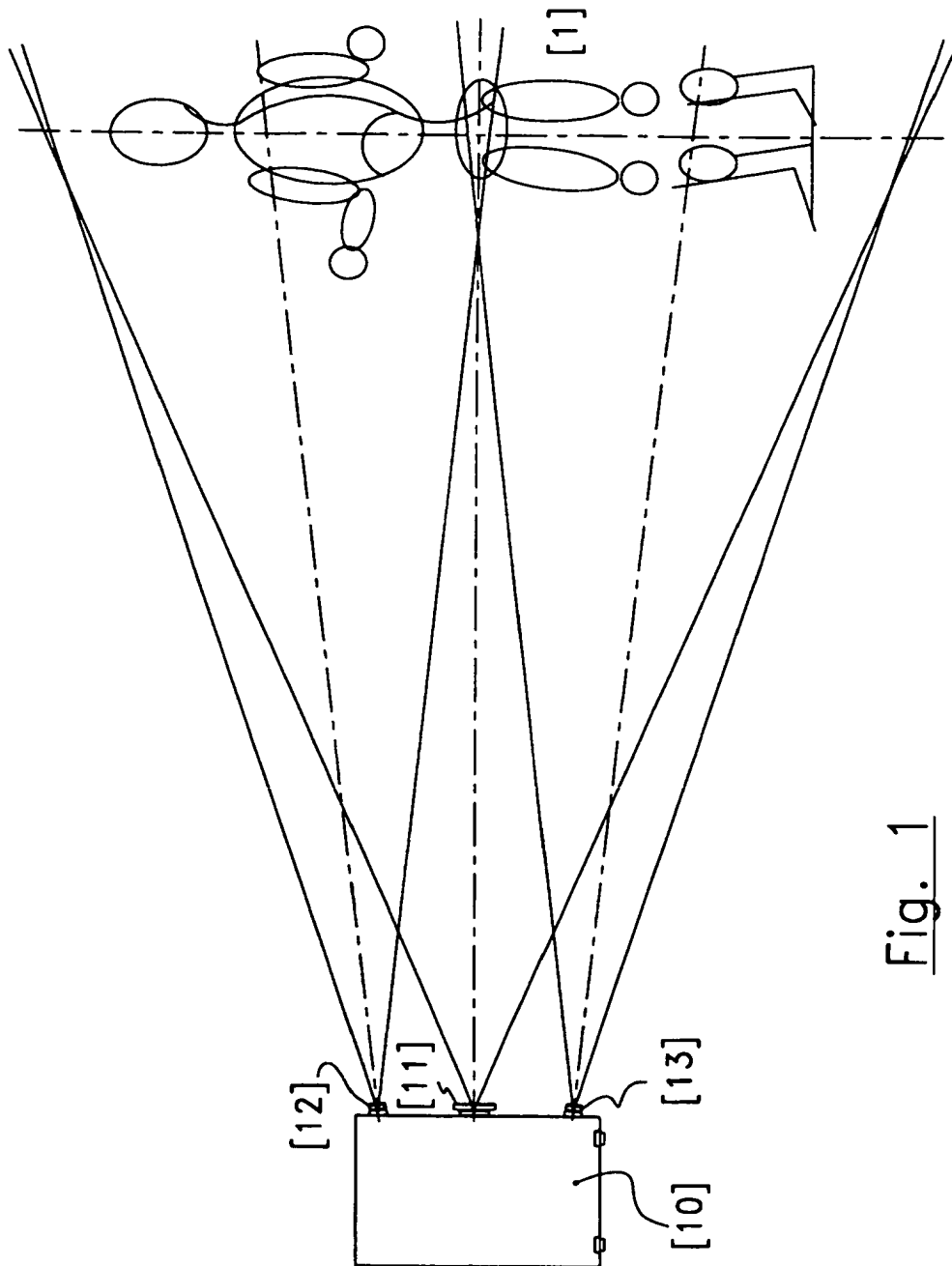
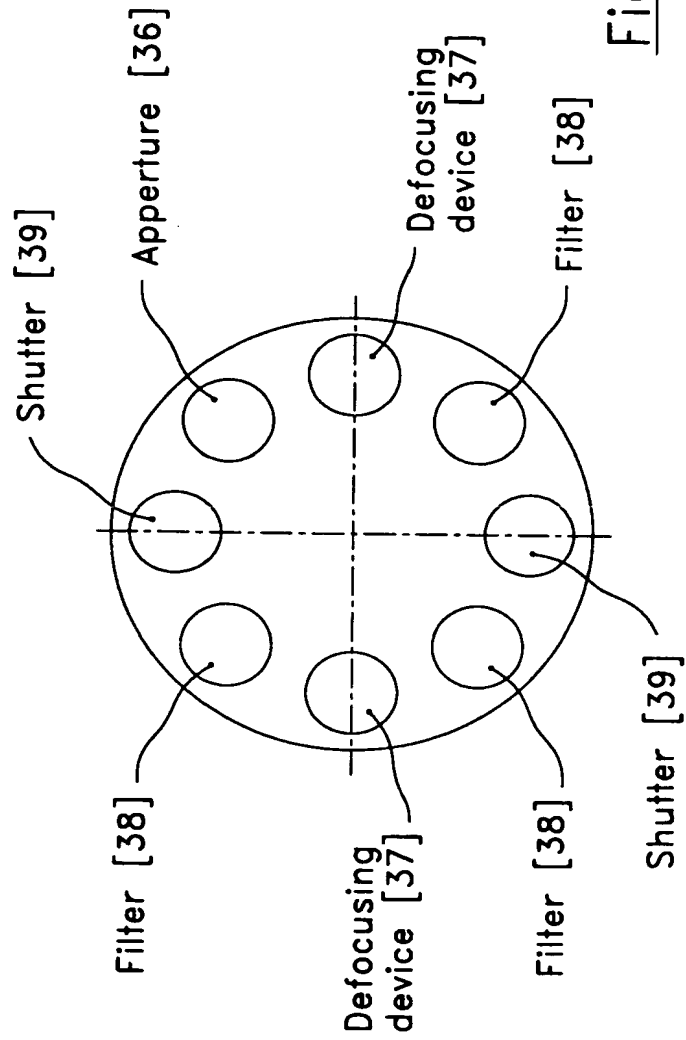


Fig. 1



Fig. 2

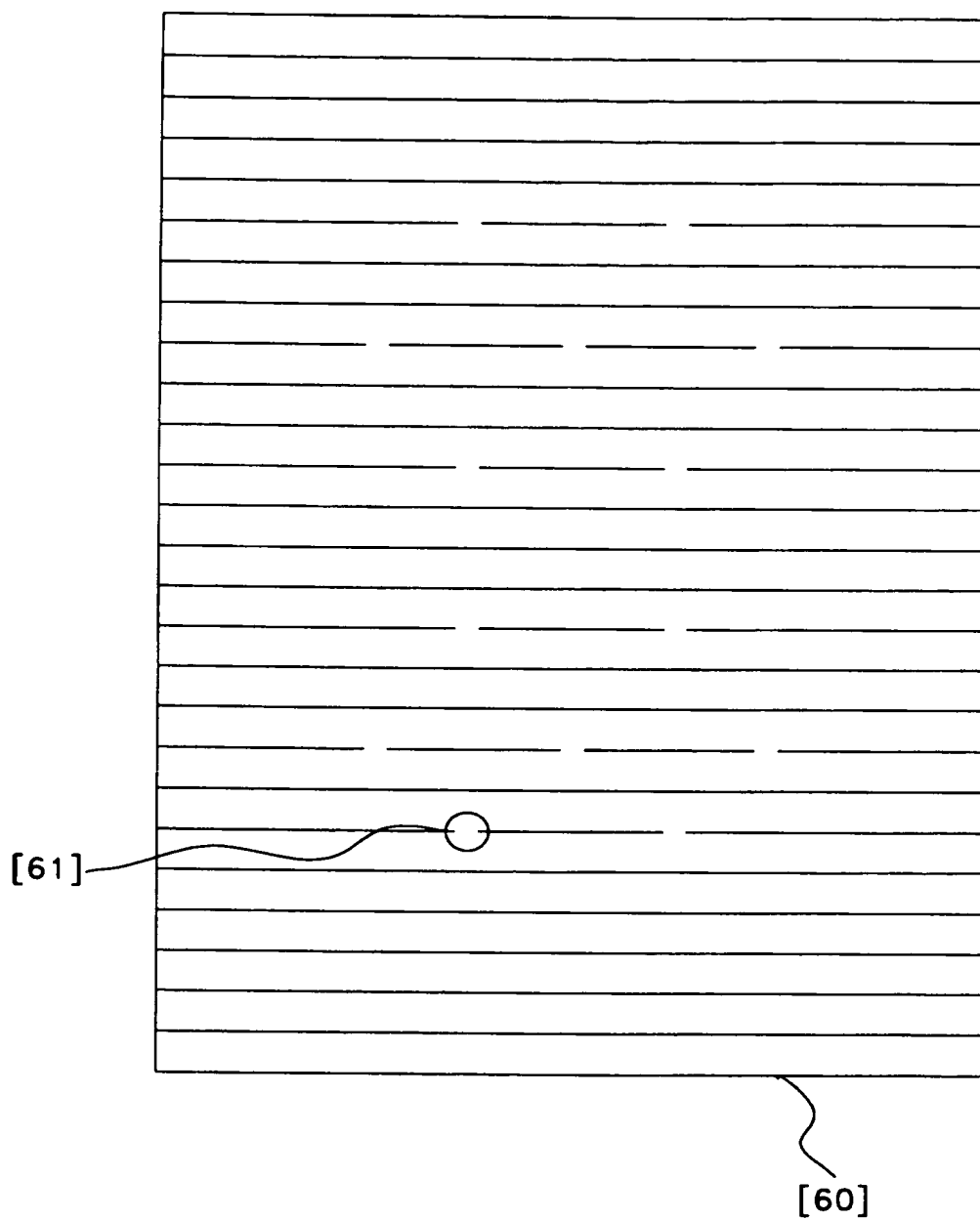


Fig. 3

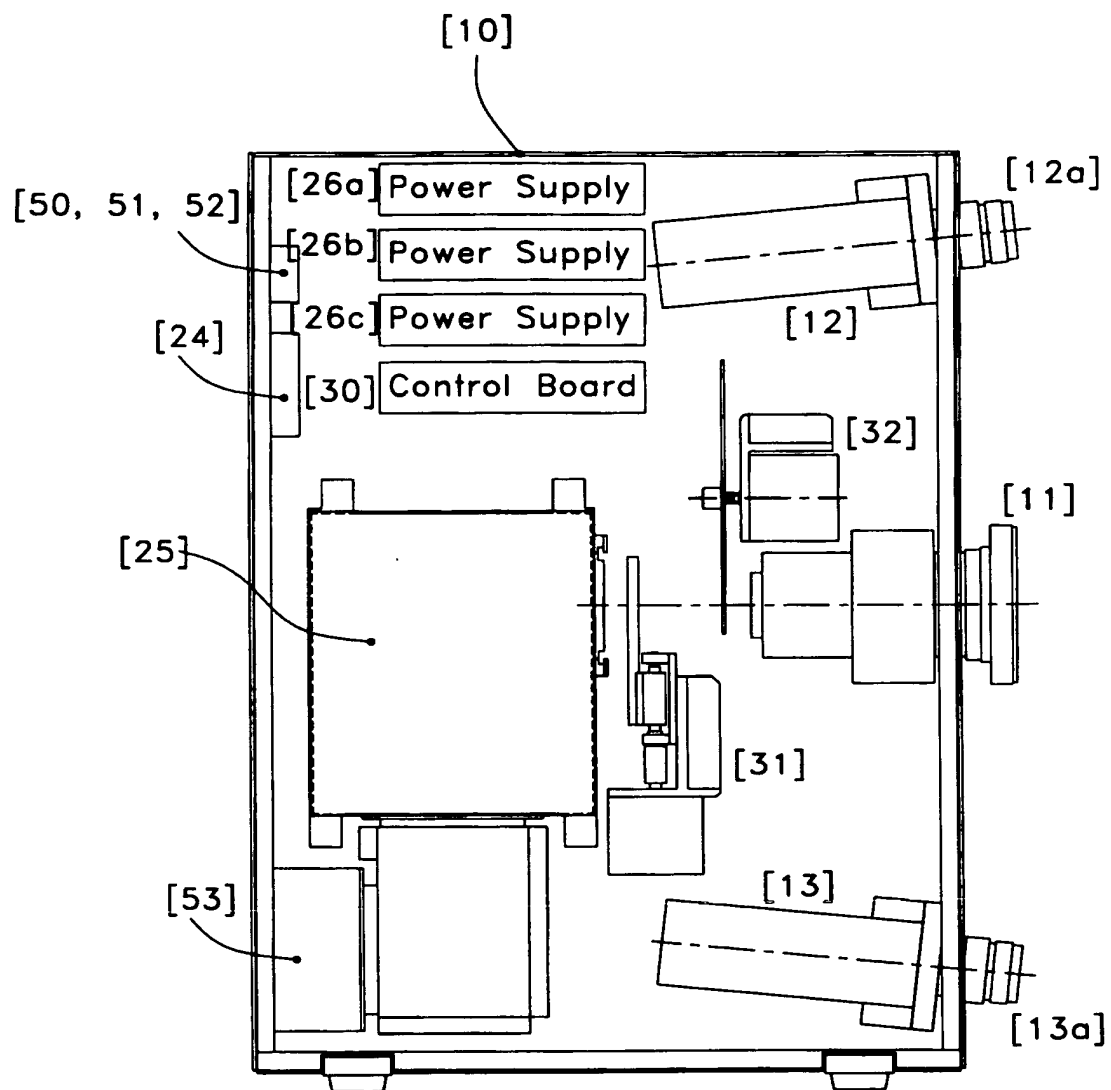


Fig. 4

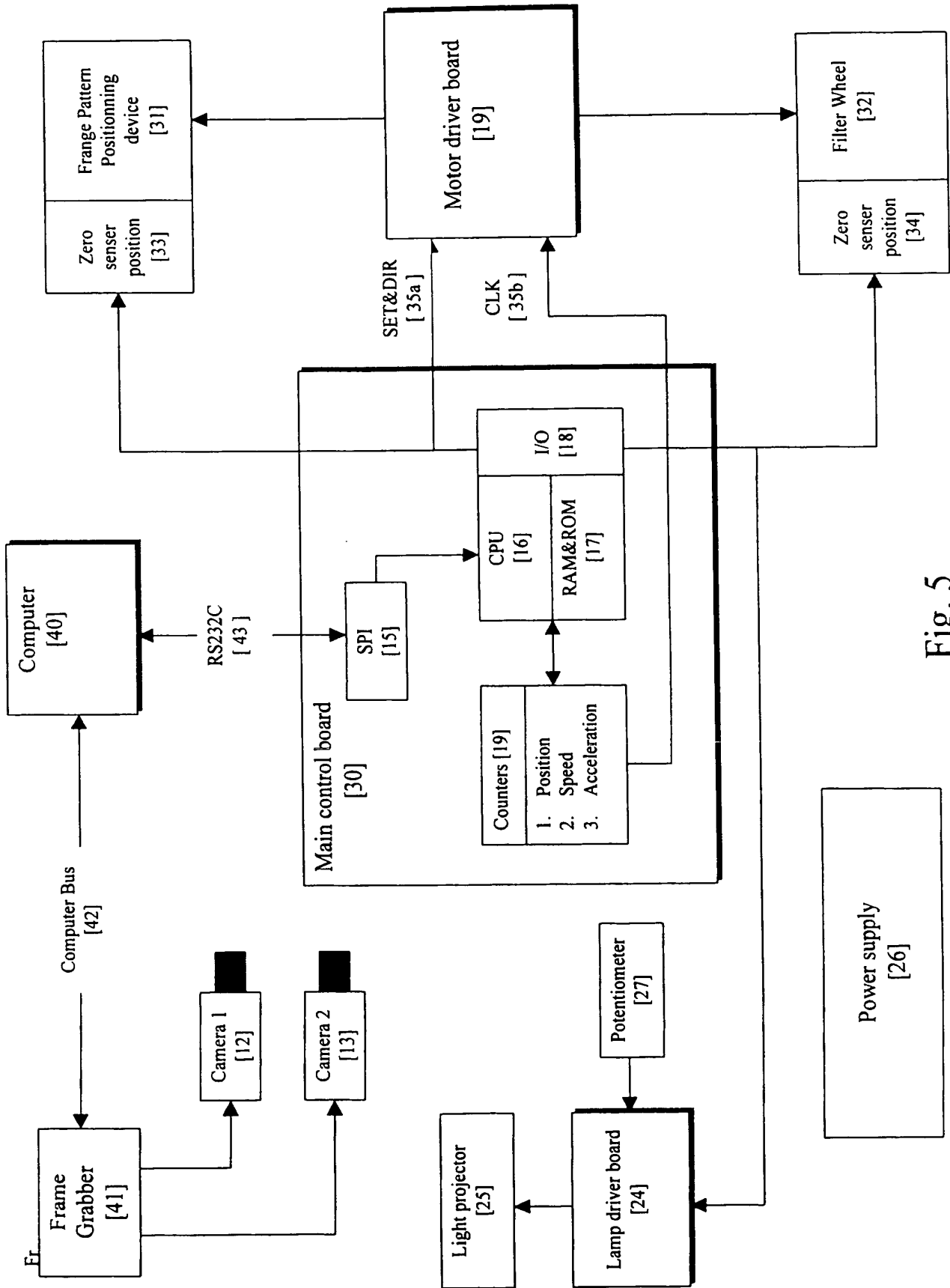
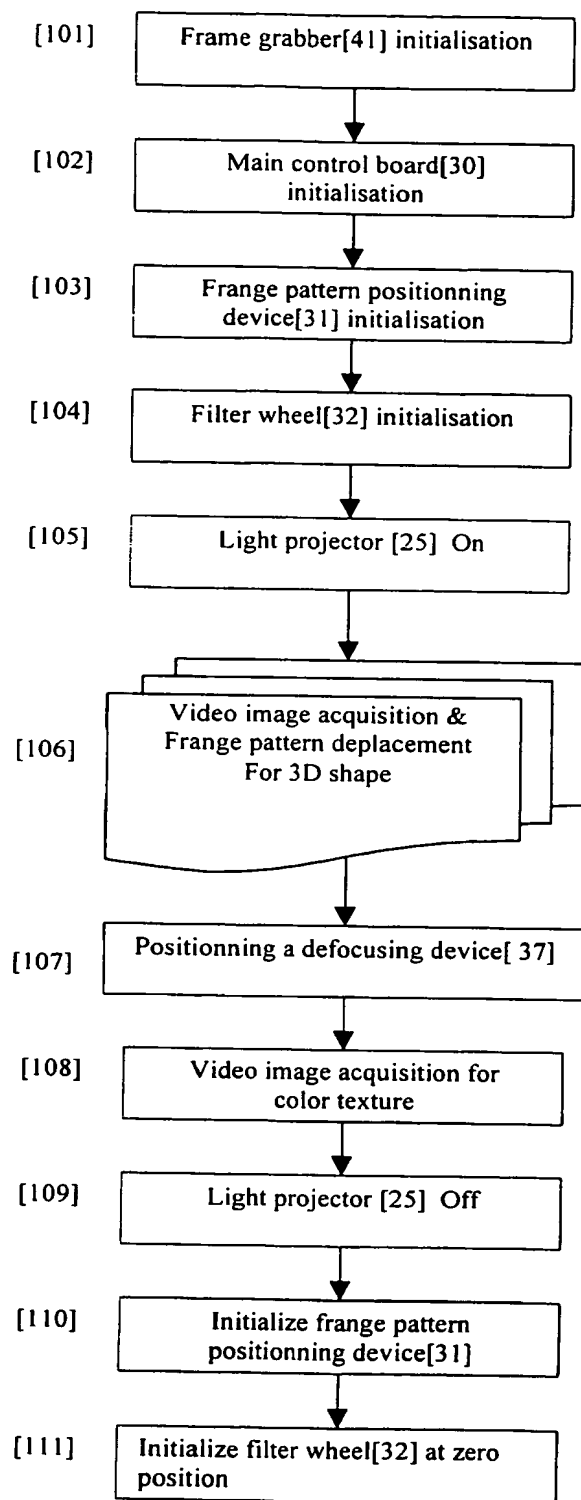


Fig. 5

Fig. 6

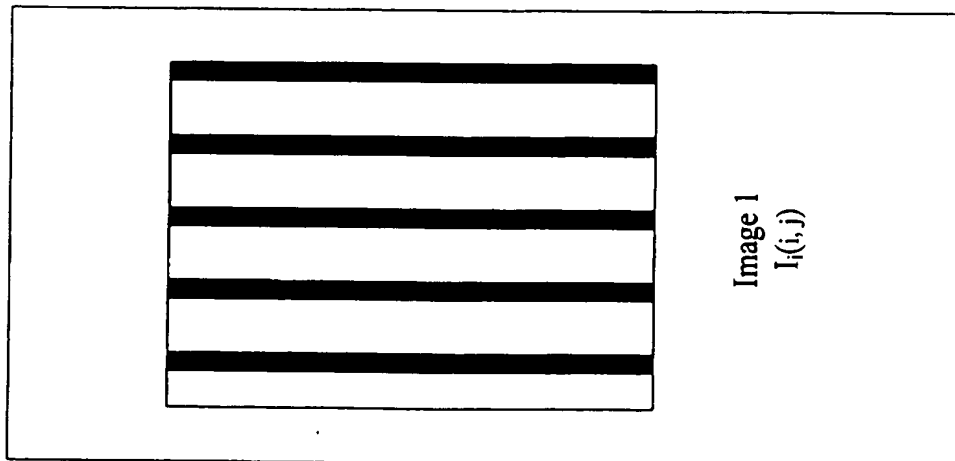


Fig. 7A

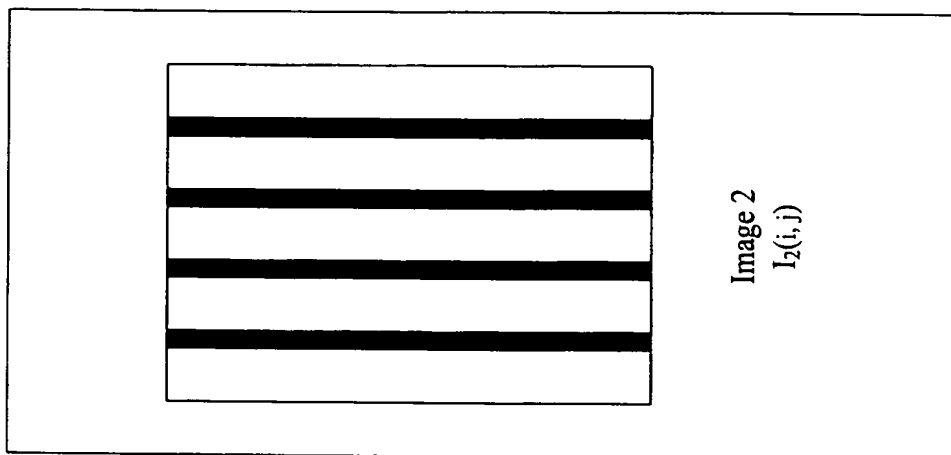


Fig. 7B

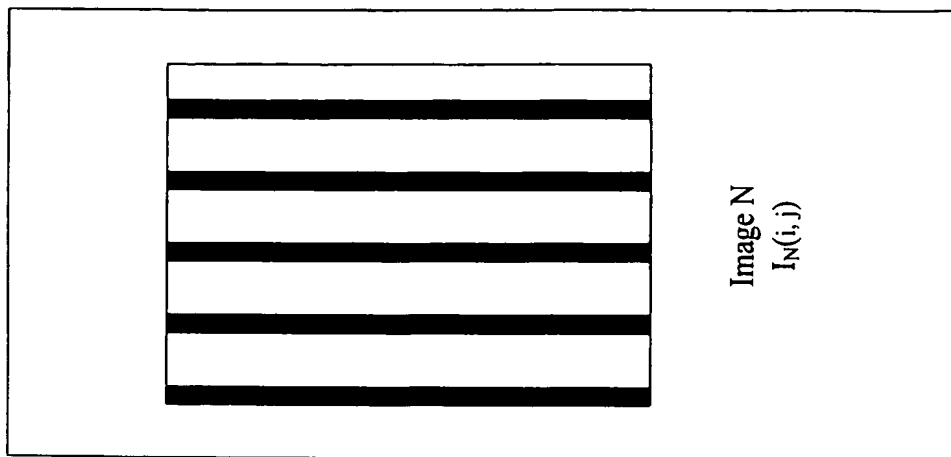


Fig. 7C